

AD-A046 973

CASE WESTERN RESERVE UNIV CLEVELAND OHIO DEPT OF MA--ETC F/G 12/1
ADAPTATION OF THE SERVICE CAPACITY IN A QUEUEING SYSTEMS WHICH --ETC(U)
NOV 77 M YADIN, S ZACKS

N00014-75-C-0529

NL

UNCLASSIFIED

TR-30

| OF |
ADA046973

12
12
12

END
DATE
FILMED
12-77

DDC

AD A 0 46973

6 ADAPTATION OF THE SERVICE CAPACITY IN A QUEUEING SYSTEMS
WHICH IS SUBJECTED TO A CHANGE IN THE ARRIVAL RATE AT UNKNOWN EPOCH.

by

10 M. Yadin and S. Zacks

14 TR-30

9 Technical Report No. 30

11 1 November 1977

12 27 p.

PREPARED UNDER CONTRACT

15 NR 00014-75-C-0529, PROJECT NR 042-276

OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for
any Purpose for the United States Government.

Department of Mathematics and Statistics
CASE WESTERN RESERVE UNIVERSITY
Cleveland, Ohio

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

403 901

DDC
RECEIVED
NOV 18 1977
B

0. Introduction

It is frequently the case that service facilities, designed to accommodate in an optimal fashion certain demand, become inadequate due to unanticipated increase in the arrival intensity of customers. This is often manifested by an excessive waiting time of customers and sometimes even saturation of the systems. To avoid these undesirable phenomena the service intensity of the system should be adapted to suit the circumstances. If the nature of change in the arrival intensity is not completely known the problem is twofold. One has to make proper inference from observations on the system about the possible epochs of increase in the arrival intensity and concurrently to make optimal decisions on the proper service capacity. In the present paper we study this problem for the special case of only one possible change in the arrival intensity, which may occur at unknown time point. The model of at most one change in the arrival intensity may be considered as an over simplification of the real problem. However, our objective is to present the problem and show an approach for obtaining an optimal solution. We allow only N possible additions to the initial service capacity of the system. We assume that one can only increase the capacity. We further restrict attention to one-server Markovian queueing systems. For such systems we develop a procedure for the optimal adaptation of the capacity. This procedure is based on a cost structure which consists of the operating cost, set-up cost and holding cost. Furthermore, the optimization is carried within a Bayesian framework. We assume that the time till the change in the arrival intensity follows an exponential prior distribution. The Markovian decision procedure is accordingly a function of two state-variables: the number of customers in the system and the posterior probability that the change has already occurred. The optimality criterion is

| | |
|--|----------------|
| Action <input checked="" type="checkbox"/> | |
| Section <input type="checkbox"/> | |
| <input type="checkbox"/> | |
| ABILITY CODES | |
| Dist. | and/or SPECIAL |
| A | |

to minimize the total expected discounted cost for the entire future. Since we make decisions under uncertainty, there is the option of changing the service capacity in N steps. We provide here an algorithm for such adaptation. In Section 1 we present the problem in more specific terms. In Section 2 we study the posterior probabilities process. In Section 3 we present the formulae for the expected cost functions at the maximal capacity. On the basis of these functions we develop in Section 4 the general functional equations describing the minimal expected cost functions, under the plausible assumption that the optimal adaptation is a non-decreasing function of each one of the state variables. This assumption is not essential but helpful in simplifying the development. Moreover, in Section 5 we provide a piece-wise linear approximation, which reduces the integral equations previously developed to simpler difference equations. An iterative method for solving these equations is discussed in Section 6. In this section we provide also computational algorithm which combines the solution of the difference equations with the determination of the optimal adaptation rules. The results of the present study can be generalized to cases where the adaptation of capacity is done by adding servers to the system.

The literature on the optimal control of queueing systems consists already of a few hundred papers. Reviews of the basic problems and approaches can be found in Crabill, Gross and Magazine (1976), Sobel (1974) and Stidham and Prabhu (1974). The present study is related directly only to the previous studies of Zacks and Yadin (1970) and Yadin (1977). These specific papers can be ascribed, according to the classification of Crabill, Gross and Magazine (1976) to the general category of "service process control" of dynamic systems. We prefer, however, to call our process an adaptation process, since the service capacity cannot be decreased but only increased.

1. The basic problem

Consider a single server queueing system in which arrivals follow a Poisson process. This process starts with an arrival rate λ_0 and at an unknown time point, τ the arrival rate changes to λ_1 . We assume that λ_0 and λ_1 are known and that $0 < \lambda_0 < \lambda_1 < \infty$.

We adopt a Bayesian model, according to which the epoch τ of shift from λ_0 to λ_1 is assumed to be a random variable having a prior exponential distribution with mean $1/\alpha$. The service times of this system are conditionally independent random variables having exponential distributions. The capacity is defined as the expected number of customers that the system can serve per unit of time. The system is set initially to have a capacity μ_0 , which is related to the initial arrival rate λ_0 . It is possible, however, to increase the capacity in N steps from μ_0 to μ_N , according to specified values:

$$\mu_0 < \mu_1 < \dots < \mu_N$$

The time points at which the changes in capacity take place are called adaptation points. There is no possibility of decreasing the capacity of the system. The problem is to determine the optimal adaptation times in a way which minimizes the total expected discounted cost of operating the system. This cost consists of three elements: the cost of maintaining service at a given capacity per unit of time, $S(\mu)$; the cost of holding x customers at the system one unit of time, $Q(x)$ and the cost of increasing the capacity (set-up cost) from μ_i to μ_j ($0 \leq i < j \leq K$), $c(i,j)$.

2. Posterior distributions based on the arrival times.

Let $\Lambda(t)$ be the arrival rate at time t , $0 < t < \infty$. According to the model

$$(2.1) \quad \Lambda(t) = \lambda_0 I\{t < \tau\} + \lambda_1 I\{t \geq \tau\};$$

where $I\{A\}$ is the indicator function of t , with respect to the set A . Accordingly, $\Lambda(t)$ is a random function. We derive here the posterior distributions of $\Lambda(t)$, given the past history of the arrival process.

Let θ_1 denote the length of time elapsed since the most recent arrival till the present time t . More generally, counting the arrivals from the time t in a backward fashion, let θ_k ($k = 1, 2, \dots$) denote the length of time since the k -th arrival till the present time t . In this notation, the last k arrival time points, prior to t are $t - \theta_1, t - \theta_2, \dots, t - \theta_k$. We also set $\theta_0 \equiv 0$. Let $N(t)$ denote the number of arrivals in the time interval $(0, t]$. In the event of $\{N(t) = n\}$ we set $\theta_{n+1} \equiv t$; and the posterior probability of $\{\Lambda(t) = \lambda_0\}$, given $\mathcal{S}_t = (N(t) = n, \theta_1, \theta_2, \dots, \theta_n)$ is

$$(2.2) \quad P\{\Lambda(t) = \lambda_0 | \mathcal{S}_t\} = \frac{\alpha \int_0^t e^{-\alpha\tau} L(\tau, \mathcal{S}_t) d\tau}{\alpha \int_0^t e^{-\alpha\tau} L(\tau, \mathcal{S}_t) d\tau}$$

where $L(\tau; \mathcal{S}_t)$ is the likelihood function of τ for a given (t, \mathcal{S}_t) . This likelihood function is expressed by

$$(2.3) \quad L(\tau; \mathcal{S}_t) = \lambda_0^n e^{-\lambda_0 t} I\{\tau > t\} + \sum_{k=0}^n \lambda_0^{n-k} \lambda_1^k e^{-\lambda_0 \tau - \lambda_1 (\tau - t)} I\{t - \theta_{k+1} < \tau < t - \theta_k\}$$

Substituting (2.3) in (2.2) we obtain after proper integrations the following formula for the posterior probabilities

$$(2.4) \quad P\{\Lambda(t) = \lambda_0 | g_t\} = \frac{\lambda_1 - \lambda_0 - \alpha}{\lambda_1 - \lambda_0 - \alpha + \alpha \sum_{k=0}^n \rho^k \left[e^{-\theta_k(\lambda_1 - \lambda_0 - \alpha)} - e^{-\theta_{k+1}(\lambda_1 - \lambda_0 - \alpha)} \right]}$$

where $\rho = \lambda_1/\lambda_0$. Define the stochastic process $\{A(t), t \geq 0\}$ which, at every point of time t , is a function of the sufficient statistic g_t , as given by the formula

$$(2.5) \quad A(t) = \sum_{k=0}^n \rho^k \left[e^{-\theta_k \alpha \eta} - e^{-\theta_{k+1} \alpha \eta} \right]$$

where $\eta = (\lambda_1 - \lambda_0 - \alpha)/\alpha$. Notice that $A(0) = 0$ with probability one. Every decision process based on the behavior of the realization of the posterior probabilities process $P\{\Lambda(t) = \lambda_0 | g_t\}$ can be equivalently based on the realization of the $A(t)$ -process, since

$$(2.6) \quad P\{\Lambda(t) = \lambda_0 | g_t\} = \frac{\eta}{\eta + A(t)}, \quad 0 \leq t < \infty.$$

We investigate here some of the properties of the $A(t)$ -process. We focus attention on the case of $\eta > 0$. The case of $\eta \leq 0$ is not of interest, since it implies an early change of λ_0 . The process $A(t)$ is induced by the non-homogeneous Poisson process in the following fashion. Consider the time interval $[t, t+h]$. If there is no arrival at this time interval then

$$(2.6) \quad A(t+h) = e^{-\alpha \eta h} [A(t) - 1] + 1.$$

On the other hand, if t is an arrival point then

$$(2.7) \quad A(t+) = \rho A(t-).$$

In other words, the $A(t)$ process has discontinuities at the arrival points. At t -points which are not arrival points the process has a smooth realization with derivatives given by

$$(2.8) \quad \frac{d}{dt} A(t) = -\alpha\eta[A(t) - 1].$$

This shows that $A(t)$ is an increasing process as long as $A(t) < 1$. Once $A(t)$ jumps above the level $A = 1$ it always stays there. Moreover, after crossing the level $A = 1$ the process is strictly decreasing between arrival points, at which it jumps upwards.

Let $E_i(a)$, $i = 0, 1$, designate the conditional expectation of $A(t)$ immediately after an arrival point, given that the level of the process immediately after the previous arrival is a , and $\Lambda(t) = \lambda_i$ for all t between these two arrival points. This conditional expectation is

$$(2.9) \quad E_i(a) = \rho \int_0^{\infty} \lambda_i e^{-\lambda_i t} \left\{ e^{-\alpha\eta t} [a - 1] + 1 \right\} dt$$

$$= \frac{\rho}{\lambda_i + \alpha\eta} (\lambda_i a + \alpha\eta), \quad i = 0, 1.$$

Since we consider here cases of $\eta > 0$ one could immediately verify that $E_i(a) > a$, $i = 0, 1$. Thus, the imbedded sequence of $A(t+)$ at the arrival points is a submartingale which diverges a.s. to infinity.

This in conjunction with (2.6) implies that the posterior probability $P\{\Lambda(t) = \lambda_0 | \mathcal{G}_t\} \rightarrow 0$ a.s., as $t \rightarrow \infty$, even if there is no change in λ_0 . This is not surprising, since the prior probability $P\{\Lambda(t) = \lambda_0\} \rightarrow 0$ exponentially fast, as $t \rightarrow \infty$.

3. The Expected Cost Functions at Maximal Service Capacity.

Due to the Markovian structure of the queueing process, the total expected discounted cost for the entire future, at epoch t , given that $\Lambda(t) = \lambda_i$ ($i = 0, 1$); the system operates at capacity μ_j ($j = 1, \dots, N$) and there are $X(t) = x$ customers in the queue is independent of t . We denote this cost function, therefore, by $M_{ij}(x)$. In the present section we present the expressions for $M_{1N}(x)$ and $M_{0N}(x)$.

3.1 Derivation of $M_{1N}(x)$.

Let $\mu_j(x) = \mu_j$ if $x \geq 1$ and $\mu_j(0) = 0$. The change in the queue size, x , is induced by two independent Poisson processes of rates λ_1 and $\mu_j(x)$. Accordingly, the time until the next change in x is exponentially distributed, with parameter $\lambda_1 + \mu_j(x)$. At the time of change the queue increases from x to $x+1$ with probability $\lambda_1/(\lambda_1 + \mu_j(x))$ and decreases to $x-1$ with probability $\mu_j(x)/(\lambda_1 + \mu_j(x))$. Thus, $M_{1N}(x)$ satisfies the functional equation

$$(3.1) \quad M_{1N}(x) = \int_0^{\infty} (\lambda_1 + \mu_N(x)) \cdot \exp[-(\lambda_1 + \mu_N(x))\theta] \cdot \left\{ \int_0^{\theta} e^{-p\tau} [S(\mu_N) + Q(x)] d\tau + e^{-p\theta} \left[\frac{\lambda_1}{\lambda_1 + \mu_N(x)} M_{1N}(x+1) + \frac{\mu_N(x)}{\lambda_1 + \mu_N(x)} M_{1N}(x-1) \right] \right\} d\theta.$$

Notice that e^{-Pr} , $0 < p < \infty$, is the present worth factor. Let

$$(3.2) \quad H(\mu, x) = \frac{1}{p} [S(\mu) + Q(x)].$$

Notice that $H(\mu, x)$ is the total discounted cost of maintaining a service rate μ and queue size x for the entire future. Substituting $H(\mu, x)$ in (3.1) we obtain the functional equation

$$(3.3) \quad M_{1N}(x) = \frac{\lambda_1 M_{1N}(x+1) + \mu_N(x) M_{1N}(x-1) + pH(\mu_N, x)}{\lambda_1 + \mu_N(x) + p}$$

The solution of this functional equation can be approximated for any $H(\mu, x)$ function. However, in special cases one could get the solution in a closed form. Zacks and Yadin (1970) derived a closed form expression for $M_{1N}(x)$, for the case of $H(\mu, x) = \frac{1}{p} [S(\mu) + qx]$. The explicit solution obtained is

$$(3.4) \quad M_{1N}(x) = \frac{1}{p} \left\{ S(\mu_N) + q \left[x + \frac{\lambda_1 - \mu_N}{p} + \frac{(\xi_{1N}(p))^{x+1}}{1 - \xi_{1N}(p)} \right] \right\}$$

where

$$(3.5) \quad \xi_{1j}(p) = \frac{1}{2\lambda_1} \left[\lambda_1 + \mu_j + p - ((\lambda_1 + \mu_j + p)^2 - 4\lambda_1 \mu_j)^{1/2} \right],$$

$i = 0, 1, \dots; j = 1, \dots, N.$

3.2 Derivation of $M_{ON}(x)$.

The functional equation for $M_{ON}(x)$ is derived as in the previous case of $M_{1N}(x)$, taking into account that the arrival rate λ_0 may change to λ_1 before the change in the queue size. Accordingly, we obtain the equation

$$(3.6) \quad M_{ON}(x) = \frac{\lambda_0 M_{ON}(x+1) + \mu_N(x) M_{ON}(x-1) + (\alpha + p) H^*(\mu_N, x)}{\lambda_0 + \mu_N(x) + \alpha + p},$$

where

$$(3.7) \quad H^*(\mu_N, x) = \frac{1}{\alpha+p} [\alpha M_{1N}(x) + pH(\mu_N, x)].$$

For the special case of $H(\mu, x) = \frac{1}{p} [S(\mu) + qx]$ we obtain, according to (3.5) and (3.7)

$$(3.8) \quad H^*(\mu_N, x) = \frac{1}{p} \left[S(\mu_N) + \frac{q\alpha(\lambda_1 - \mu_N)}{p(\alpha+p)} \right] + \frac{q}{p} x + \frac{\xi_{1N}^{x+1}(p)}{p(\alpha+p)(1-\xi_{1N}(p))}.$$

The function $H^*(\mu_N, x)$ can be written as the sum of the two functions

$$(3.9) \quad H^{(1)}(\mu_N, x) = \frac{1}{p} \left[S(\mu_N) + \frac{\alpha}{p} \frac{q(\lambda_1 - \mu_N)}{(\alpha+p)} + qx \right]$$

and

$$(3.10) \quad H^{(2)}(\mu_N, x) = \frac{q}{p} \cdot \frac{\alpha}{\alpha+p} \cdot \frac{\xi_{1N}(p)}{1-\xi_{1N}(p)} \cdot \xi_{1N}^x(p).$$

By substituting the function $H^{(1)}(\mu_N, x)$, for $H^*(\mu_N, x)$ in (3.6), we obtain a functional equation similar to (3.3). The solution of this equation is given therefore, in analogy with (3.4), by

$$(3.11) \quad M_{ON}^{(1)}(x) = \frac{1}{p} \left[S(\mu_N) + q \left(x + \frac{\lambda^* - \mu_N}{p} + \frac{\xi_{ON}^{x+1}(\alpha+p)}{1-\xi_{ON}(\alpha+p)} \right) \right],$$

where

$$(3.12) \quad \lambda^* = \frac{\alpha \lambda_1 + p \lambda_0}{\alpha+p}.$$

When $H^{(2)}(\mu_N, x)$ is substituted in (3.6) the solution assumes the form

$$(3.13) \quad M_{ON}^{(2)}(x) = \frac{q\alpha}{p} \cdot \frac{\xi_{1N}(p)}{1 - \xi_{1N}(p)} G_x(\xi_{1N}(p), \alpha+p)$$

$$= \frac{q\alpha\xi_{1N}(p) \left[\frac{\xi_{1N}^{x+1}(p)}{1 - \xi_{1N}(p)} - \frac{\xi_{ON}^{x+1}(\alpha+p)}{1 - \xi_{ON}(\alpha+p)} \right]}{p \lambda_0 [\xi_{1N}(p) - \xi_{ON}(\alpha+p)] [\psi_{ON}(\alpha+p) - \xi_{1N}(p)]},$$

where $G_x(z, s)$, $|z| \leq 1$, $0 < s < \infty$ is the double transform of the transition distribution of the $M/M/1$ queueing process with λ_0 and μ_N . Moreover, $\psi_{ON}(\alpha+p) = \mu_N/\lambda_0 \xi_{ON}(\alpha+p)$. The denominator of (3.17) can be further simplified and the formula of $M_{ON}^{(2)}(x)$ becomes

$$(3.14) \quad M_{ON}^{(2)}(x) = \frac{q}{p} B \left[\frac{\xi_{1N}^{x+1}(p)}{1 - \xi_{1N}(p)} - \frac{\xi_{ON}^{x+1}(\alpha+p)}{1 - \xi_{ON}(\alpha+p)} \right]$$

where

$$(3.15) \quad B = \frac{\alpha \xi_{1N}(p)}{\lambda_0 (\xi_{1N}(p) - \xi_{ON}(\alpha+p)) (\psi_{ON}(\alpha+p) - \xi_{1N}(p))}$$

$$= \alpha / [\alpha + (\lambda_0 - \lambda_1)(1 - \xi_{1N}(p))].$$

Finally, since equation (3.6) is linear its solution of (3.10) is the sum of $M_{ON}^{(1)}(x)$ and $M_{ON}^{(2)}(x)$. Accordingly,

$$(3.16) \quad M_{ON}(x) = \frac{1}{p} \left\{ S(\mu_N) + q \left[x + \frac{\lambda^* - \mu_N}{p} \right. \right.$$

$$\left. \left. + (1 - B) \frac{\xi_{ON}^{x+1}(\alpha+p)}{1 - \xi_{ON}(\alpha+p)} + B \frac{\xi_{1N}^{x+1}(p)}{1 - \xi_{1N}(p)} \right] \right\}$$

In the special case of $\lambda_0 = \lambda_1$ we have $B = 1$ and (3.16) is reduced to (3.4).

4. Minimal Expected Cost Functions For Monotone Policies

In the present section we provide the recursive equations on which the optimization process is based. The decision concerning the optimal service capacity depends on the current capacity, μ_n ($n = 0, 1, 2, \dots, N$); on the number of customers in the station, x , and on the posterior probability $\Pi(t) = P\{\Lambda(t) = \lambda_1 | S_t\}$.

We start with the supposition that the adaptation policy should be monotone in the queue size, $X(t)$, for a fixed value of the posterior probability $\Pi(t)$ and monotone in $\Pi(t)$ for a fixed $X(t)$. In other words, if for a given value of $\Pi(t)$ it is optimal, at time t , to shift from μ_n to μ_j ($j > n$); the optimal decision for all $x > X(t)$ is to shift from μ_n to μ_k ($k \geq j > n$). Similarly, for a given value of $X(t)$, and since $\Pi(t)$ is an increasing function of $A(t)$, an optimal shift from μ_n to μ_j ($j > n$) under $A(t)$ implies that the optimal shift for all $a > A(t)$ is from μ_n to μ_k ($k \geq j > n$). We further impose the restriction that no change in the service capacity will take place as long as $A(t) < 1$. It can then be shown that under these conditions the optimal decision epochs are those at which arrivals occur. Due to the Markovian properties of the decision process, increase in capacity should take place either when the queue size increases or when the posterior probability of λ_1 increases. Since the posterior probability of λ_1 is an increasing function of $A(t)$ and, when $A(t) \geq 1$, it increases only at the arrival epochs. The epochs of increase in the queue size and in the posterior probability coincide.

Let $K_n(x, a)$ be the minimal (conditional) expected discounted cost at some epoch t , given that $X(t) = x$, $A(t) = a$ and the capacity is μ_n . Furthermore, let $K_{in}(x, a)$ ($i = 0, 1$) denote the conditional expected

discounted cost, given μ_n , $X(t) = x$, $A(t) = a$ and $\Lambda(t) = \lambda_1$, associated with the optimal policy. For each n , and a

$$(4.1) \quad K_n(x, a) = (a K_{in}(x, a) + \eta K_{On}(x, a)) / (\eta + a).$$

Let $I_{nj}(x, a)$ denote an indicator function, which assumes the value 1 if, under the optimal policy, an increase from μ_n to μ_j ($j = n, \dots, N$) should take place at epoch t for which $X(t) = x$ and $A(t) = a$; and assumes the value zero otherwise $\sum_{j=n}^N I_{nj}(x, a) = 1$. With the aid of the indicator functions $I_{nj}(x, a)$ we determine simultaneously with $K_{in}(x, a)$ associated functions $M_{in}(x, a)$ ($i = 0, 1$) satisfying the equations:

$$(4.2) \quad K_{in}(x, a) = \sum_{j=n}^N I_{nj}(x, a) [c(n, j) + M_{1j}(x, a)],$$

$i = 0, 1; n = 1, \dots, N$ and

$$(4.3) \quad M_{On}(x, a) = \frac{p H(\mu_n, x)}{\lambda_0 + \mu_n(x) + p + \alpha} + \int_0^{\infty} e^{-(\lambda_0 + \mu_n(x) + p + \alpha)\tau} \left[\lambda_0 K_{0,n}(x+1, \rho e^{-\alpha\eta\tau}(a-1+\rho) + \mu_n(x) M_{0,n}(x-1, e^{-\alpha\eta\tau}(a-1) + 1) + \alpha M_{1,n}(x, e^{-\alpha\eta\tau}(a-1) + 1) \right] d\tau,$$

and

$$(4.4) \quad M_{1n}(x, a) = \frac{p H(\mu_n, x)}{\lambda_1 + \mu_n(x) + p} + \int_0^{\infty} e^{-(\lambda_1 + \mu_n(x) + p)\tau} \left[\lambda_1 K_{1,n}(x+1, \rho e^{-\alpha\eta\tau}(a-1) + \rho) + \mu_n(x) M_{1n}(x-1, e^{-\alpha\eta\tau}(a-1) + 1) \right] d\tau.$$

Having determined the functions $M_{in}(x, a)$, we consider their expectation

$$(4.5) \quad M_n(x, a) = (aM_{1n}(x, a) + \eta M_{0n}(x, a)) / (\eta + a).$$

Equations (4.1), (4.2) and (4.5) imply

$$(4.6) \quad K_n(x, a) = \sum_{j=n}^N I_{nj}(x, a) [c(n, j) + M_j(x, a)].$$

It follows that

$$(4.7) \quad K_n(x, a) = \min_{n \leq j \leq N} \{c(n, j) + M_j(x, a)\}.$$

We remark that for $n = N$

$$(4.8) \quad K_N(x, a) = M_N(x, a) = (aM_{1N}(x) + \eta M_{0N}(x)) / (a + \eta).$$

Equations (4.3) and (4.4) can be expressed in terms of the M-functions only, by substituting the K-functions according to (4.2). Accordingly,

$$(4.9) \quad M_{in}(x, a) = \frac{p}{w_{in}(x)} H(\mu_n, x) + \int_0^{\infty} e^{-w_{in}(x)\tau} \left\{ \lambda_i \sum_{j=n}^N I_{nj}(x+1, \rho(a-1)e^{-\alpha\eta\tau} + \rho) [c(n, j) + M_{ij}(x+1, \rho(a-1)e^{-\alpha\eta\tau} + \rho)] + \mu_n(x) M_{in}(x-1, (a-1)e^{-\alpha\eta\tau} + 1) + \alpha M_{1n}(x, (a-1)e^{-\alpha\eta\tau} + 1) \right\} d\tau;$$

where $w_{in}(x) = \lambda_i + \mu_n(x) + \alpha + p$ ($i = 0, 1$). Notice that in (4.9), τ designates the random time of change in the queue size; capacity adaptation is done only at epochs of increase in the queue size and $A(t) = a$ changes according to the formulae given in Section 2. Making the transformation $u = (a-1)e^{-\alpha\eta\tau} + 1$, (4.9) is reduced to

$$(4.10) \quad M_{in}(x, a) = \frac{p}{w_{in}(x)} H(\mu_n, x) +$$

$$\frac{(a-1)^{-w_{in}(x)/\alpha\eta}}{\alpha\eta} \int_1^a (u-1)^{w_{in}(x)/\alpha\eta-1} \left\{ \lambda_1 \sum_{j=n}^N I_{nj}(x+1, \rho u) \cdot \right.$$

$$[c(n, j) + M_{ij}(x+1, \rho u)] + \mu_n(x) M_{in}(x-1, u) +$$

$$\left. \alpha M_{in}(x, u) \right\} du.$$

Partition the interval $(1, a)$ to the sets

$$(4.11) \quad \Theta_{nj}(x) = \{u; I_{nj}(x, \rho u) = 1\}, \quad j = n, \dots, N.$$

Since we consider only monotone policies, these sets are subintervals.

The functional equations are then, for each $i = 0, 1$;

$$(4.12) \quad M_{in}(x, a) = \frac{p}{w_{in}(x)} H(\mu_n, x) +$$

$$\frac{(a-1)^{-w_{in}(x)/\alpha\eta}}{\alpha\eta} \sum_{j=n}^N \int_{\Theta_{nj}(x+1)}^a (u-1)^{w_{in}(x)/\alpha\eta-1} \left\{ \lambda_1 [c(n, j) + M_{ij}(x+1, \rho u)] \right.$$

$$\left. + \mu_n(x) M_{in}(x-1, u) + \alpha M_{in}(x, u) \right\} du.$$

We introduce now boundary functions for optimal policies, $a_{nj}(x)$, $j = n, \dots, N$, $x = 0, 1, \dots$ such that, if the value of $A(t)$ lies between $a_{nj}(x)$ and $a_{n, j+1}(x)$ it is optimal to increase the capacity from μ_n to μ_j . We formally define $a_{nn}(x) = \rho$ and $a_{n, N+1}(x) = \infty$, for each $n = 1, \dots, N$ and $x = 0, 1, \dots$. The connection between the subintervals $\Theta_{nj}(x)$ and the boundary functions $a_{nj}(x)$ is specified by

$$(4.13) \quad \Theta_{nj}(x) = \left\{ u; \frac{a_{nj}(x)}{\rho} \leq u < \frac{a_{n, j+1}(x)}{\rho} \right\}.$$

Notice that for all j such that $a < \frac{a_{nj}(x)}{\rho}$ the subinterval $\Theta_{nj}(x)$ is empty. Accordingly, define the index

$$(4.14) \quad j_n(x, a) = \max \left\{ j; n \leq j, a \geq \frac{a_{nj}(x)}{\rho} \right\}$$

and set $a_{n, j_n(x, a)+1}(x) = \rho a$.

5. Piecewise Linear Approximation

In the present section we derive a linear functional equation which approximates (4.12) on a denumerable set of points. More specifically, consider a partition of the range of a according to the logarithmic scale $r = \rho^{1/\ell}$, where ℓ is a predetermined integer. Furthermore, consider sequences of integer boundary points, $d_{nj}(x)$, in the new scale such that

$$(5.1) \quad a_{nj}(x) = r^{d_{nj}(x)}, \quad j \geq n.$$

For $i = 0, 1$, let

$$(5.2) \quad G_{in}(x, a) = \frac{1}{w_{in}(x)} \left\{ \lambda_i K_{in}(x+1, \rho a) + \mu_n(x) M_{in}(x-1, a) + \alpha M_{in}(x, a) \right\},$$

In these terms, functions $M_{in}(x, a)$ at the points $a = r^k$, $k = 1, 2, \dots$, are given by

$$(5.3) \quad M_{in}(x, r^k) = \frac{\rho}{w_{in}(x)} H(\mu_n, x) + \frac{w_{in}(x)}{\alpha \ell} (r^{k-1})^{-w_{in}(x)/\alpha \ell} \sum_{v=1}^k \int_{r^{v-1}}^{r^v} (u-1)^{\frac{w_{in}(x)}{\alpha \ell} - 1} G_{in}(x, u) du$$

For each $v = 1, \dots, k$, define

$$(5.4) \quad S_{in}(x, v) = \frac{w_{in}(x)}{\alpha\eta} \int_{r^{v-1}}^{r^v} (u-1) \frac{w_{in}(x)}{\alpha\eta} - 1 G_{in}(x, u) du,$$

and write

$$(5.5) \quad M_{in}(x, r^k) = \frac{p}{w_{in}(x)} H(\mu_n, x) + (r^k - 1)^{-w_{in}(x)/\alpha\eta} \sum_{v=1}^k S_{in}(x, v).$$

By simple substitution we can verify that the functions $M_{in}(x, r^k)$ can be determined recursively, for each $k \geq 1$, according to

$$(5.6) \quad M_{in}(x, r^k) = \left[1 - \frac{r^{k-1} - 1}{r^k - 1} \frac{w_{in}(x)/\alpha\eta}{} \right] \cdot \frac{p}{w_{in}(x)} H(\mu_n, x) + \left(\frac{r^{k-1} - 1}{r^k - 1} \right) \frac{w_{in}(x)/\alpha\eta}{} M_{in}(x, r^{k-1}) + (r^k - 1)^{-w_{in}(x)/\alpha\eta} S_{in}(x, k),$$

where, for $k = 0$,

$$(5.7) \quad M_{in}(x, 1) = \frac{1}{w_{in}(x)} \left\{ p H(\mu_n, x) + \lambda_i K_{in}(x+1, \rho) + \mu_n(x) M_{in}(x-1, 1) + \alpha M_{in}(x, 1) \right\}.$$

This expression is obtained directly from (4.12) by applying the L'Hospital rule. We approximate now the functions $M_{ij}(x, a)$ piecewise linearly on the intervals (r^{v-1}, r^v) $v = 1, 2, \dots, k$. The value of l which determines the logarithmic scale determines the accuracy of the approximation. The functions $G_{ij}(x, a)$ are, as in (5.2), linear combinations of $M_{in}(x, a)$

functions within these intervals. Thus, they can be represented within each interval $[r^{v-1}, r^v]$, $v = 0, 1, \dots$ by the linear functions

$$(5.8) \quad G_{in}^*(x, a) = A_{i,n}^{(v)}(x) + B_{in}^{(v)}(x)a, \quad r^{v-1} \leq a \leq r^v,$$

where

$$(5.9) \quad B_{in}^{(v)}(x) = \frac{1}{r^v - r^{v-1}} \left[G_{in}^*(x, r^v) - G_{in}^*(x, r^{v-1}) \right].$$

If we substitute $G_{in}^*(x, a)$ in (5.4) we obtain the approximation

$$(5.10) \quad S_{in}^*(x, k) = (r^k - 1)^{w_{in}(x)/\alpha\eta} G_{in}^*(x, r^k) - (r^{k-1} - 1)^{w_{in}(x)/\alpha\eta} G_{in}^*(x, r^{k-1}) - B_{in}^{(k)}(x) \frac{\alpha\eta}{w_{in}(x) + \alpha\eta} \left[(r^k - 1) \frac{w_{in}(x)}{\alpha\eta} + 1 - (r^{k-1} - 1) \frac{w_{in}(x)}{\alpha\eta} + 1 \right].$$

Substituting (5.9) and (5.10) in (5.6) we obtain a recursive function for the approximating function $M_{in}^*(x, r^k)$, $k \geq 1$, where $M_{in}^*(x, 1) = M_{in}(x, 1)$.

6. Computing Algorithm For Recursive Solution

We provide a computing algorithm for the solution of the functional equations for $M_{in}^*(x, r^k)$. The approach is to determine first a reasonable initial solution and improve it by a sequence of iterations. It is desirable here to express the equations for $i = 0$ and $i = 1$ separately, since these equations are derived from two separate sources (4.3) and (4.4), respectively. They were combined to one expression in (4.9) for the sake of uniform treatment. Following the general backward induction scheme of dynamic programming we start with the functions corresponding to μ_N . We compute first $M_{1N}(0)$ according to (3.4) and $M_{0N}(0)$ according to (3.16). Then, for every $x = 1, 2, \dots$ we compute

$$(6.1) \quad \begin{aligned} M_{1N}(x) &= M_{1N}(x-1) + \frac{q}{p} (1 - \xi_{1N}^x(p)), \\ M_{0N}(x) &= M_{0N}(x-1) + \frac{q}{p} \left[1 - ((1-B)\xi_{0N}^x(\alpha+p) + B\xi_{1N}^x(p)) \right] \end{aligned}$$

and by (4.8) we determine their expected value $M_N(x, r^k)$. For $n = N-1$, we start with the initial solution,

$$(6.2) \quad \begin{aligned} d_{N-1, N}^{(0)}(x) &= 1 \\ K_{i, N-1}^{(0)}(x, r^k) &= c(N-1, N) + M_{iN}(x) \end{aligned}$$

for $x = 1, 2, \dots, k = \ell, \ell+1, \dots$. This initial solution corresponds to the case in which one increases the capacity following the first arrival. For this reason we set $d_{N-1, N}^{(0)}(x) = \ell$. Clearly, this is not necessarily an optimal policy. The following iterations are designed to improve the policy and simultaneously to adjust the expected cost functions. Let $M_{i, N-1}^{(v)}(x, r^k)$ denote the solutions at the v -th iteration; $v = 1, 2, \dots$. These functions are determined for $k = 1$ recursively according to the formulae

$$\begin{aligned}
 (6.3) \quad M_{1, N-1}^{(v)}(0, 1) &= \frac{1}{\lambda_1 + p} \left\{ \lambda_1 K_{1, N-1}^{(v-1)}(1, r^\ell) + S(\mu_{N-1}) \right\}, \\
 M_{0, N-1}^{(v)}(0, 1) &= \frac{1}{\lambda_0 + \alpha + p} \left\{ \lambda_0 K_{0, N-1}^{(v-1)}(1, r^\ell) + \alpha M_{1, N-1}^{(v)}(0, 1) + S(\mu_{N-1}) \right\}, \\
 M_{1, N-1}^{(v)}(x, 1) &= \frac{1}{\lambda_1 + \mu_{N-1} + p} \left\{ \lambda_1 K_{1, N-1}^{(v-1)}(x+1, r^\ell) + \right. \\
 &\quad \left. \mu_{N-1} M_{1, N-1}^{(v)}(x-1, 1) + p H(\mu_{N-1}, x) \right\} \\
 M_{0, N-1}^{(v)}(x, 1) &= \frac{1}{\lambda_0 + \mu_{N-1} + \alpha + p} \left\{ \lambda_0 K_{0, N-1}^{(v-1)}(x+1, r^\ell) + \right. \\
 &\quad \left. \mu_{N-1} M_{0, N-1}^{(v)}(x-1, 1) + \alpha M_{1, N-1}^{(v)}(x, 1) + p H(\mu_{N-1}, x) \right\}.
 \end{aligned}$$

To obtain $M_{i, N-1}^{(v)}(x, r^k)$ for $k \geq 1$, we determine first

$$\begin{aligned}
 G_{1, N-1}^{(v)}(0, r^k) &= \frac{\lambda_1}{\lambda_1 + p} K_{1, N-1}^{(v-1)}(r, r^{k+\ell}) \\
 (6.4) \quad G_{0, N-1}^{(v)}(0, r^k) &= \frac{1}{\lambda_0 + \alpha + p} \left\{ \lambda_0 K_{0, N-1}^{(v-1)}(1, r^{k+\ell}) + \alpha M_{1, N-1}^{(v)}(0, r^k) \right\}
 \end{aligned}$$

$$G_{1,N-1}^{(v)}(x, r^k) = \frac{1}{\lambda_1 + \mu_{N-1} + p} \left\{ \lambda_1 K_{1,N-1}^{(v-1)}(x+1, r^{k+1}) + \mu_{N-1} M_{1,N-1}^{(v)}(x-1, r^k) \right\},$$

and

$$G_{0,N-1}^{(v)}(x, r^k) = \frac{1}{\lambda_0 + \mu_{N-1} + p + \alpha} \left\{ \lambda_0 K_{0,N-1}^{(v-1)}(x+1, r^{k+1}) + \mu_{N-1} M_{0,N-1}^{(v)}(x-1, r^k) + \alpha M_{1,N-1}^{(v)}(x, r^k) \right\},$$

For the sake of convenience, define

$$(6.5) \quad w_{i,N-1}^*(x) = \begin{cases} \lambda_0 + \mu_{N-1}(x) + \alpha + p, & \text{if } i = 0 \\ \lambda_1 + \mu_{N-1}(x) + p, & \text{if } i = 1; \end{cases}$$

and

$$(6.6) \quad A_{i,N-1}^*(x, r^k) = \left[\frac{r^{k-1} - 1}{r^k - 1} \right]^{w_{i,N-1}^*(x)/\alpha\eta}$$

$$B_{i,N-1}^*(x, r^k) = 1 - \frac{\alpha\eta}{w_{i,N-1}^*(x) + \alpha\eta} \left[1 - A_{i,N-1}^*(x, r^k) \frac{r^{k-1} - 1}{r^k - r^{k-1}} \right]$$

The functions $M_{i,N-1}^{(v)}(x, r^k)$ are given then by

$$(6.7) \quad M_{i,N-1}^{(v)}(x, r^k) = A_{i,N-1}^*(x, r^k) M_{i,N-1}^{(v)}(x, r^{k-1})$$

$$+ (1 - A_{i,N-1}^*(x, r^k)) \frac{p}{w_{in}^*(x)} H(\mu_{N-1}, x) +$$

$$\left[1 - A_{i,N-1}^*(x, r^k) - B_{i,N-1}^*(x, r^k) \right] G_{i,N-1}^{(v)}(x, r^{k-1})$$

$$+ B_{i,N-1}^*(x, r^k) G_{i,N-1}^{(v)}(x, r^k).$$

Following these computations we determine at the v -th iteration the posterior expectation

$$(6.8) \quad M_{N-1}^{(v)}(x, r^k) = \frac{1}{\eta + r^k} \left[\eta M_{0, N-1}^{(v)}(x, r^k) + r^k M_{1, N-1}^{(v)}(x, r^k) \right].$$

Furthermore, we correct the K functions and determine the boundary points for the present iteration. Thus, we set

$$(6.9) \quad K_{N-1}^{(v)}(x, r^k) = \min \left\{ M_{N-1}^{(v)}(x, r^k), K_{N-1}^{(v-1)}(x, r^k) \right\}$$

$$d_{N-1, N}^{(v)}(x) = \min \left\{ k; K_{N-1}^{(v)}(x, r^k) < M_{N-1}^{(v)}(x, r^k) \right\}.$$

Let

$$(6.10) \quad K_{i, N-1}^{(v)}(x, r^k) = \begin{cases} M_{i, N-1}^{(v)}(x, r^k) & , \text{ if } k < d_{N-1, N}^{(v)}(x) \\ K_{i, N-1}^{(v-1)}(x, r^k) & , \text{ if } k \geq d_{N-1, N}^{(v)}(x). \end{cases}$$

For a given $K_{i, N-1}^{(v-1)}(x, r^k)$ the functions $M_{i, N-1}^{(v)}(x, r^k)$ are determined exactly by the recursive equations developed above. At the final stage

of the v -th iteration the function $K_{i, N-1}^{(v-1)}(x, r^k)$ is changed to

$K_{i, N-1}^{(v)}(x, r^k)$ which is the basis for the $(v+1)$ st iteration. For each x and k , the sequence $\left\{ K_{N-1}^{(v)}(x, r^k); v \geq 0 \right\}$ is non-increasing and

therefore converges to a limit as $v \rightarrow \infty$. Moreover, the corresponding

sequence $\left\{ d_{N-1, N}^{(v)}(x); v \geq 0 \right\}$ is non-decreasing. This is verified by

observing that for each x , if $k \geq d_{N-1, N}^{(v)}(x)$ then $K_{N-1}^{(v-1)}(x, r^k) =$

$K_{N-1}^{(0)}(x, r^k)$. Thus, if $d_{N-1, N}^{(v)}(x) < d_{N-1, N}^{(v-1)}(x)$ then for all k ,

$d_{N-1, N}^{(v)}(x) < k \leq d_{N-1, N}^{(v-1)}(x)$, $K_{N-1}^{(v)}(x, r^k) = K_{N-1}^{(0)}(x, r^k)$. On the other hand,

at the $(v-1)$ st iteration with $k' = d_{N-1, N}^{(v-1)}(x)$, $K_{N-1}^{(v-1)}(x, r^{k'}) < K_{N-1}^{(0)}(x, r^{k'})$.

But this contradicts the fact that $K_{N-1}^{(v)}(x, r^k) \leq K_{N-1}^{(v-1)}(x, r^k)$ at every

x and k . Under the following condition on the cost functions

$$(6.11) \quad p c(N-1, N) + (S(\mu_{N-1}) - S(\mu_N)) + \frac{q}{p} (\mu_{N-1} - \mu_N) < 0$$

there exists a sufficiently large value of x , \bar{x} say, so that for all $x \geq \bar{x}$ it is optimal to switch to μ_N even if $\Lambda(t) = \lambda_0$, for all t . We therefore restrict our computations only for $x \leq \bar{x}$. Similarly, we restrict the computations for $k \leq \bar{k}$; where \bar{k} is sufficiently large so that the posterior probability that $\Lambda(t) = \lambda_1$ is at least $1 - \epsilon$ ($\epsilon > 0$ arbitrarily small) for all $k > \bar{k}$. On the restricted domain of $x \leq \bar{x}$ and $k \leq \bar{k}$, the iterations converge in a finite number of steps. For $n < N-1$ the procedure is very similar. However, to determine $K_{i,n}^{(0)}(x, r^k)$ we set

$$(6.12) \quad \begin{aligned} K_n^{(0)}(x, r^k) &= \min_{j > n} \{c(n, j) + M_j(x, r^k)\}, \\ d_{nj}^{(0)}(x) &= \min \left\{ k; K_n^{(0)}(x, r^k) = c(n, j) + M_j(x, r^k) \right\}; \quad j = n, \dots, N \\ I_{nj}^{(0)}(x, k) &= \begin{cases} 1, & \text{if } d_{nj}^{(0)}(x) \leq k < d_{n, j+1}^{(0)}(x) \\ 0, & \text{if otherwise} \end{cases}, \quad j = n, \dots, N. \end{aligned}$$

and

$$(6.13) \quad K_{in}^{(0)}(x, r^k) = \sum_{j=n+1}^N I_{nj}^{(0)}(x, k) [c(n, j) + M_{ij}(x, r^k)].$$

The functions $M_{ij}(x, r^k)$ and $M_j(x, r^k)$ are the ones obtained by the iterations in the previous stages ($j > m$). The rest of the computations are similar. In formula (6.9) we should replace $d_{N-1, N}^{(v)}(x)$ by

$$(6.14) \quad d_n^{(v)}(x) = \min \left\{ k; K_n^{(v)}(x, r^k) < M_n^{(v)}(x, r^k) \right\}$$

and

$$d_{nj}^{(v)}(x) = \max \left\{ d_{nj}^{(v-1)}(x), d_n^{(v)}(x) \right\}, \quad j = n, \dots, N.$$

where $d_{nj}^{(0)}(x)$ is specified in (6.12) Formula (6.10) is now replaced by

$$K_{i,n}^{(v)}(x, r^k) = \begin{cases} M_{i,n}^{(v)}(x, r^k) & , \text{ if } k < d_n^{(v)}(x) \\ K_{i,n}^{(v-1)}(x, r^k) & , \text{ if } k \geq d_n^{(v)}(x). \end{cases}$$

7. References.

- [1] Crabill, T. B., Gross, D. and M. J. Magazine (1976).
"A Catalog of Research on Optimal Design and Control of Queues,"
Serial T-341, Program In Logistics, The George Washington University.
- [2] Sobel, M. J. (1974).
"Optimal Operation of Queues," Proceedings of a Conference at Western
Michigan University, May 10-12, 1973. Mat
Mathematical Methods In Queueing Theory, Lecture Notes in Economics
and Mathematical Systems, 231-261, Springer-Verlag, Berlin.
- [3] Stidham, S. Jr., and N. U. Prabhu (1974).
"Optimal Control of Queueing Systems," Proceedings of a Conference
at Western Michigan University, May 10-12, 1973.
Mathematical Methods In Queueing Theory, Lecture Notes in Economics
and Mathematical Systems, 263-294, Springer-Verlag, Berlin.
- [4] Yadin, M. (1977).
Markovian Queueing Systems With Stochastically Varying Arrival Rates,
Technical Report No. 29, ONR Project NR 042-276, Department of Mathe-
matics and Statistics, Case Western Reserve University.
- [5] Zacks, S. and M. Yadin (1970).
Analytic characterization of the optimal control of a queueing system,
J. Appl. Prob. 7: 617-613.

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|--|-----------------------|--|
| 1. REPORT NUMBER Technical Report No. 30 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) Adaptation of the Service Capacity in a Queueing Systems which is Subjected to a Change in the Arrival Rate at Unknown Epoch. | | 5. TYPE OF REPORT & PERIOD COVERED |
| 7. AUTHOR(s) M. Yadin and S. Zacks | | 6. PERFORMING ORG. REPORT NUMBER |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics and Statistics Case Western Reserve University | | 8. CONTRACT OR GRANT NUMBER(s) NR 00014-75-C-0529 Project NR 042-276 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Virginia 22217 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 12. REPORT DATE November 1, 1977 |
| | | 13. NUMBER OF PAGES 24 |
| | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report) DISTRIBUTION OF DOCUMENT UNLIMITED | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number) M/M/1 Queueing Systems, Adaptation of Service Intensity, Dynamic Programming, Bayes Adaptive Policy. <i>lambda sub 1 lambda sub zero mu 1 mu 2 mu N tau</i> | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The paper studies the problem of optimal adaptation of an M/M/1 queueing station, when the arrival rate λ_0 of customers shifts at unknown epoch, τ to a known value, λ_1 . The service intensity of the system starts at μ_0 and can be increased at most N times to $\mu_1 < \mu_2 < \dots < \mu_N$. The cost structure consists of the cost of changing μ_i to μ_j ($i+1 \leq j \leq N$); of maintaining (OVER) | | |

service at rate μ (per unit of time) and of holding customers at the station (per unit of time). Adaptation policies are constrained by the fact that μ can be only increased. A Bayes solution is derived, under the prior assumption that τ has an exponential distribution. This solution minimizes the total expected discounted cost for the entire future.

τ